STAT 429 Final Report: On the Relationship Between CO2, Cattle, and Global Temperature

Jeff Massman, NetID: massman4

11/26/2022

1. Abstract

While carbon dioxide (CO2) is a well-known contributor to the warming of the Earth. However, the potential role that cattle play in this warming, as known producers of methane, is unknown. In this paper, we addressed three objectives: quantifying the relationship between CO2 levels, cattle population, and temperature anomaly; determining if the cattle population plays a role in this warming; and forecasting the temperature anomaly for the next five years. We trained two different time series regression models, and selected a single canonical model among the two, using various criteria, which includes linear and quadratic trend terms. We found that for every 1 PPM increase in CO2, there is a 0.07422^{o} C increase in the temperature anomaly. In terms of the cattle population, our results were largely inconclusive as cattle was right above the 5% significance cutoff, and we recommend further analysis. Similarly, we found that the temperature anomaly is forecasted to decrease in the next five years, which is either true (yet surprising), or an artefact of the short-term effectiveness of trend terms in the model. Ultimately, we recommend further investigation into other factors surrounding cattle, as well as more analysis into the long-term effectiveness of trend terms.

2. Introduction

In this section, we introduce the main topic of this report, including background information, motivation, and the stated objectives. We also provide some preliminary information on the data.

2.1 General Background

The fact that the climate is changing is, at least scientifically, uncontroversial; everyone's heard that 97% of climate scientists are in agreement about the changing climate. Therefore, the aim of this report is not to investigate whether or not climate change is occurring, as that is a foregone conclusion. Furthermore, it is no secret that the levels of carbon dioxide (CO2) in the atmosphere play a critical role in the warming of the Earth. What is less clear, however, is the role, if any, that cattle play in this process. After all, cattle are known to produce large amounts of methane, which is a much more potent greenhouse gas than CO2. And so the objective of this report is to accomplish the following:

1.) To quantitatively demonstrate the relationship between CO2 levels, the cattle population, and the global temperature anomaly,

2.) To determine if the cattle population is significantly correlated with the global temperature anomaly, and

3.) To use this information to forecast the temperature anomoly for the next five years.

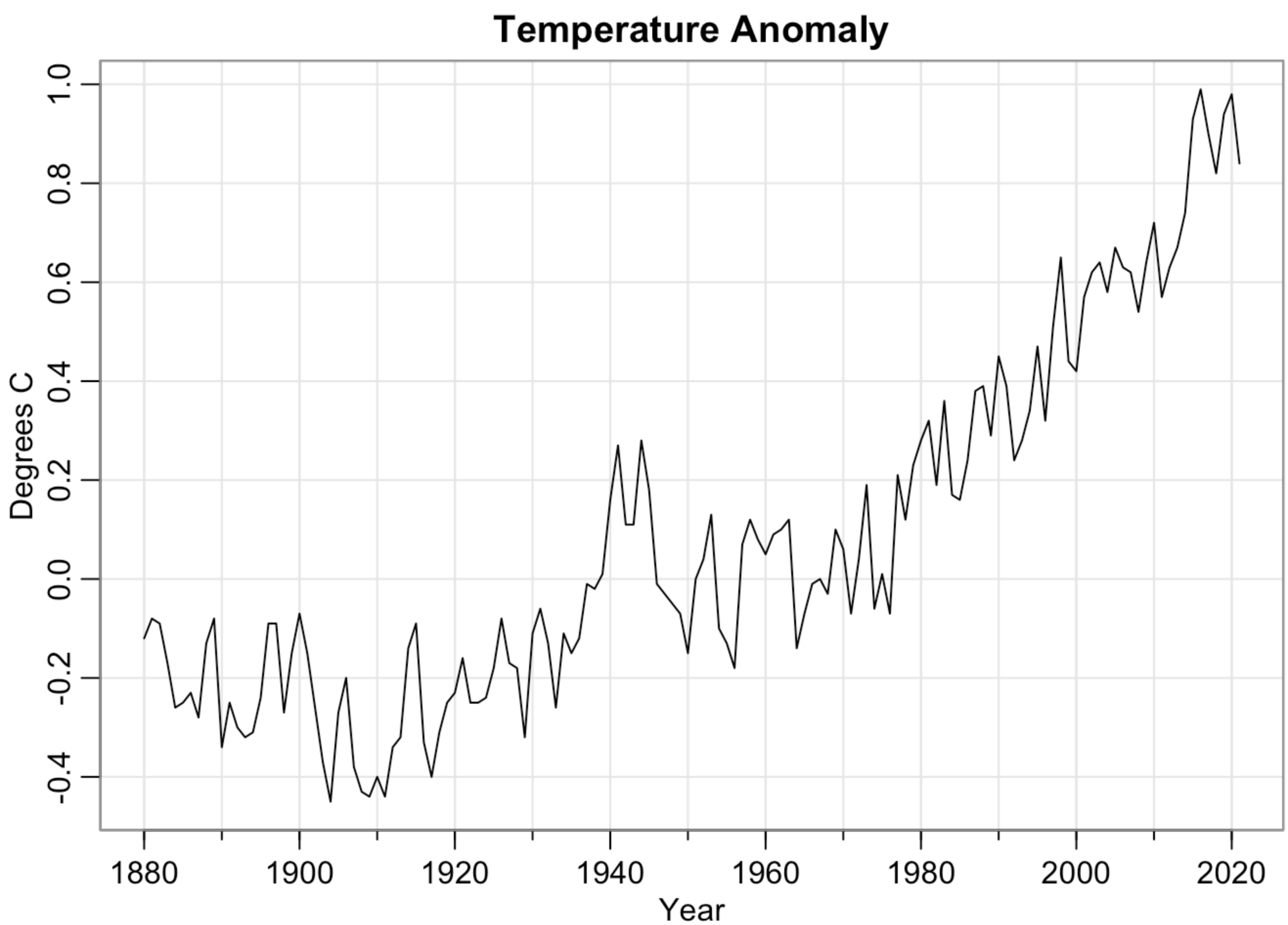
2.2 Data Background

In this section, we will provide some information on the data, including plots and commentary. For information regarding the source of the data, see section 6, the bibliography at the end of this report.

2.2.1 Temperature Anomaly Data

Temperature anomaly is defined as the deviation of the average annual global temperature from some fixed point. In our data, the fixed point is 14^{o} Celsius. The deviations are measured in degrees Celsius, and are measured in annual increments.

Below is a time series plot of the data.

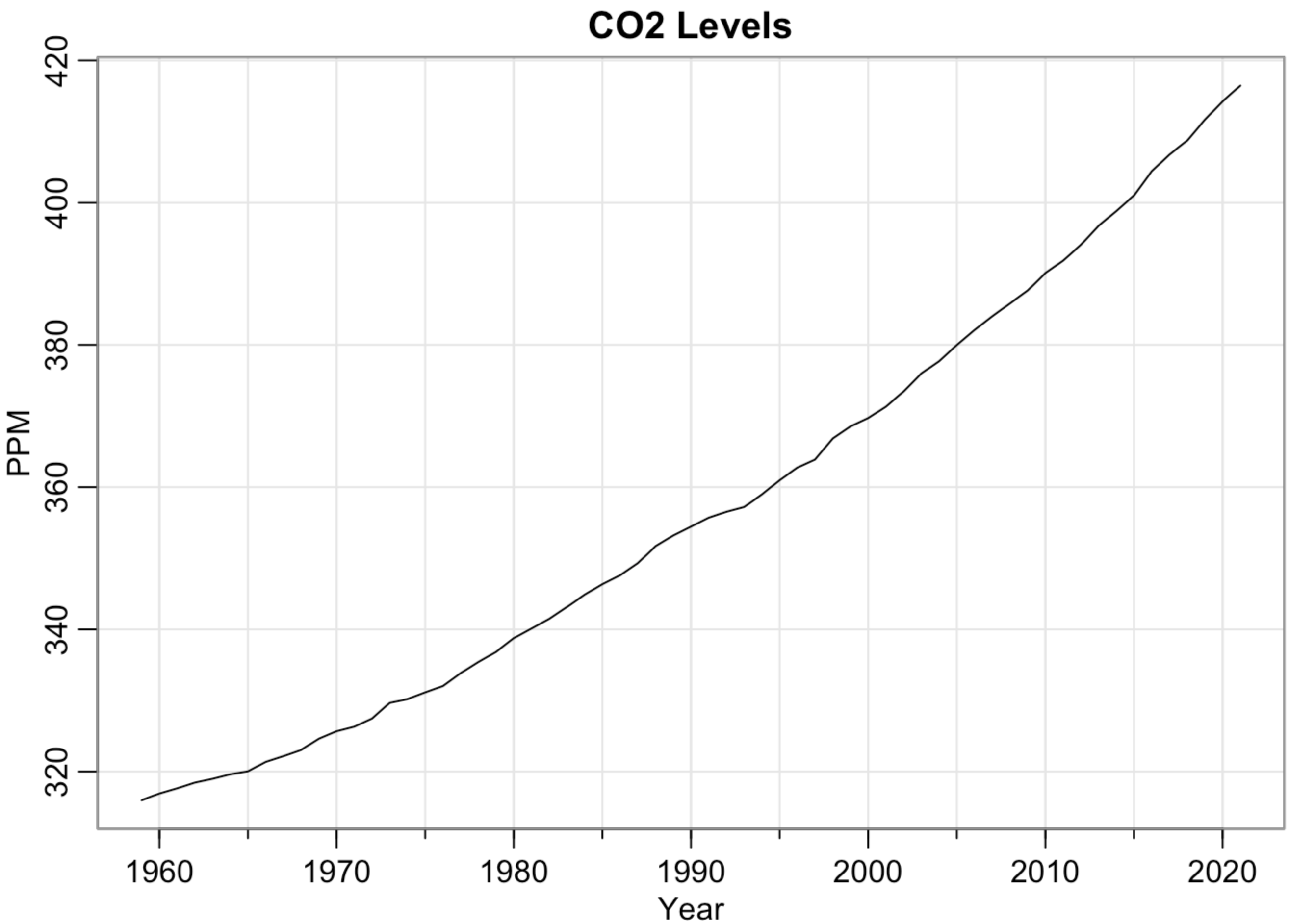


We can immediately see an upwards trend in the data after about 1910, and therefore the series is not stationary. Nevertheless the variance seems to be constant. This data starts in the year 1880, but for the purposes of this report, we will only focus on the years 1960-2021 since the cattle dataset has no records prior to 1960.

2.2.2 CO2 Data

The CO2 data measures the annual average concentration of CO2 in the atmosphere, as measured from the Mauna Loa Observatory. It is measured in parts per million (PPM), which is a metric indicating how many CO2 molecules are present in a sample of one million air particles.

Below is a time series plot of the data:

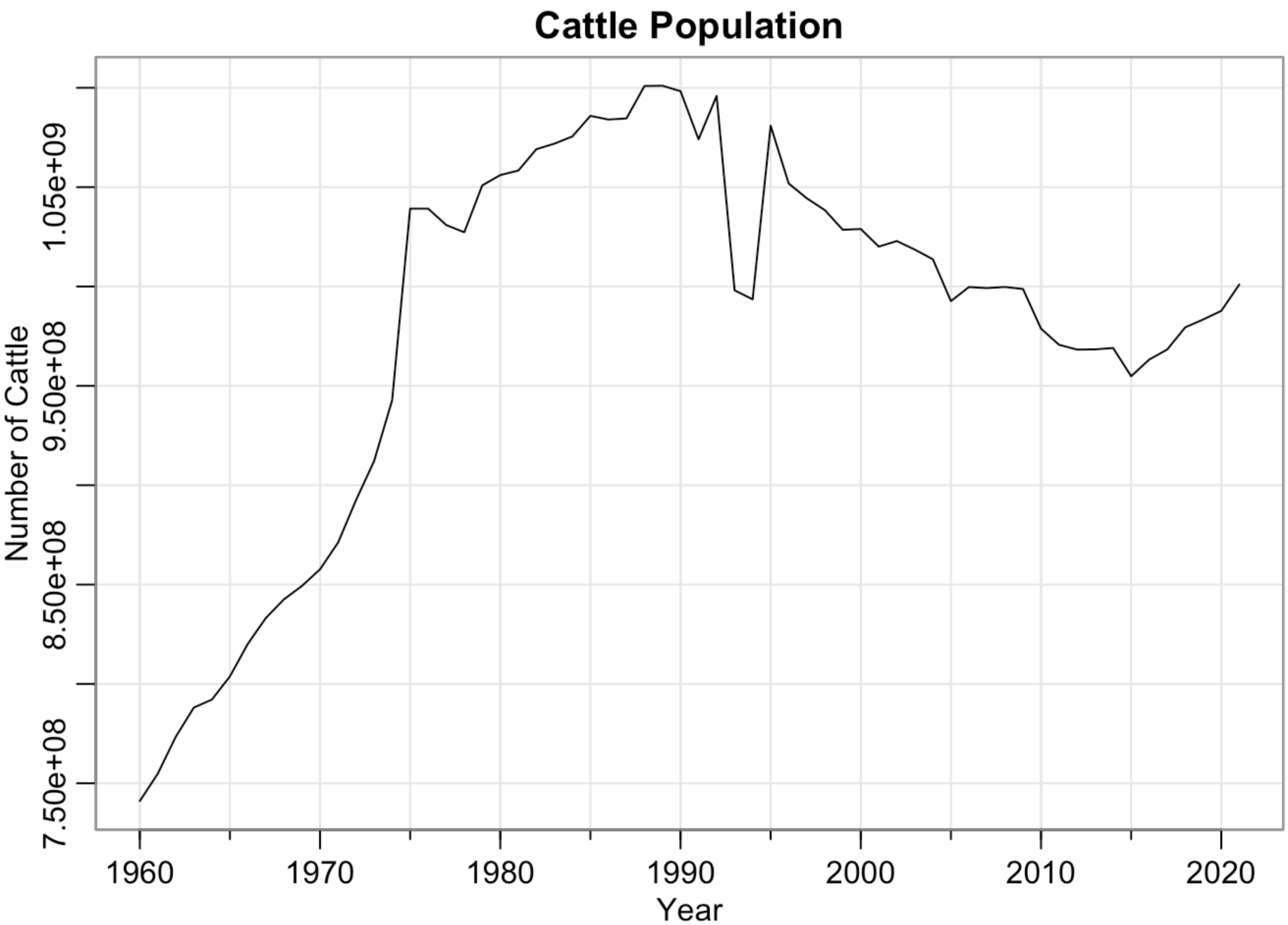


There is a clear upwards trend with very little noise. The series is clearly not stationary, but the variance is again constant throughout.

2.2.3 Cattle Data

The cattle data set measures the annual global cattle population from 1960 to 2021.

Below is a time series plot of the data:



This data has an initial upwards trend before fluctuating a bit. There is also a significant drop in 1993 and then a significant increase in 1995 before slowly dropping off. As of 2022, the population seems to be increasing again since 2015. The series is not stationary, but the variance does seem to be constant, large spikes notwithstanding.

3. Methods

In this section, we build a time series regression model with the aim of answering our three main objectives. More details are in the relevant sections below.

3.1. Notation

In this paper, we will use the following notation:

- t, the time parameter; here the unit of t is years
- G_t, the average global temperature anomoly, treated as a time series process variable
- CO_t, the CO2 level, treated as a time series process variable
- CA_t, the cattle population, treated as a time series process variable
- W_t, the usual white noise process with mean 0 and variance σ^2; note that this process is stationary and normally distributed

3.2. Modelling

In this section, we consider two different time series regression models. We then evaluate each model based on certain criteria (more details later) and choose the "better" model to serve as our canonical model throughout this report. This model will be used in section 4 to address the three objectives outlined in section 2.1.

3.2.1. Model A

The first model we will consider is the simple naive model

$$G_t = \beta_0 + \beta_1 CO_t + \beta_2 CA_t + W_t.$$

This model is somewhat reasonable to start off with as it is expresses the simplest base relationship between the variables. Below is a summary table of the model fitting results:

```
##
## Call:
## lm(formula = temp_ts$Anomaly[81:142] ~ as.numeric(co2_ts$CO2[3:64]) +
##     cattle_ts$Cattle)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.189397 -0.073868 -0.006968  0.075634  0.186035
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -3.151e+00  1.600e-01 -19.699   <2e-16 ***
## as.numeric(co2_ts$CO2[3:64])  9.898e-03  4.343e-04  22.789   <2e-16 ***
## cattle_ts$Cattle             -1.529e-11  1.333e-10  -0.115    0.909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09256 on 59 degrees of freedom
## Multiple R-squared:  0.9117, Adjusted R-squared:  0.9087
## F-statistic: 304.7 on 2 and 59 DF,  p-value: < 2.2e-16
```

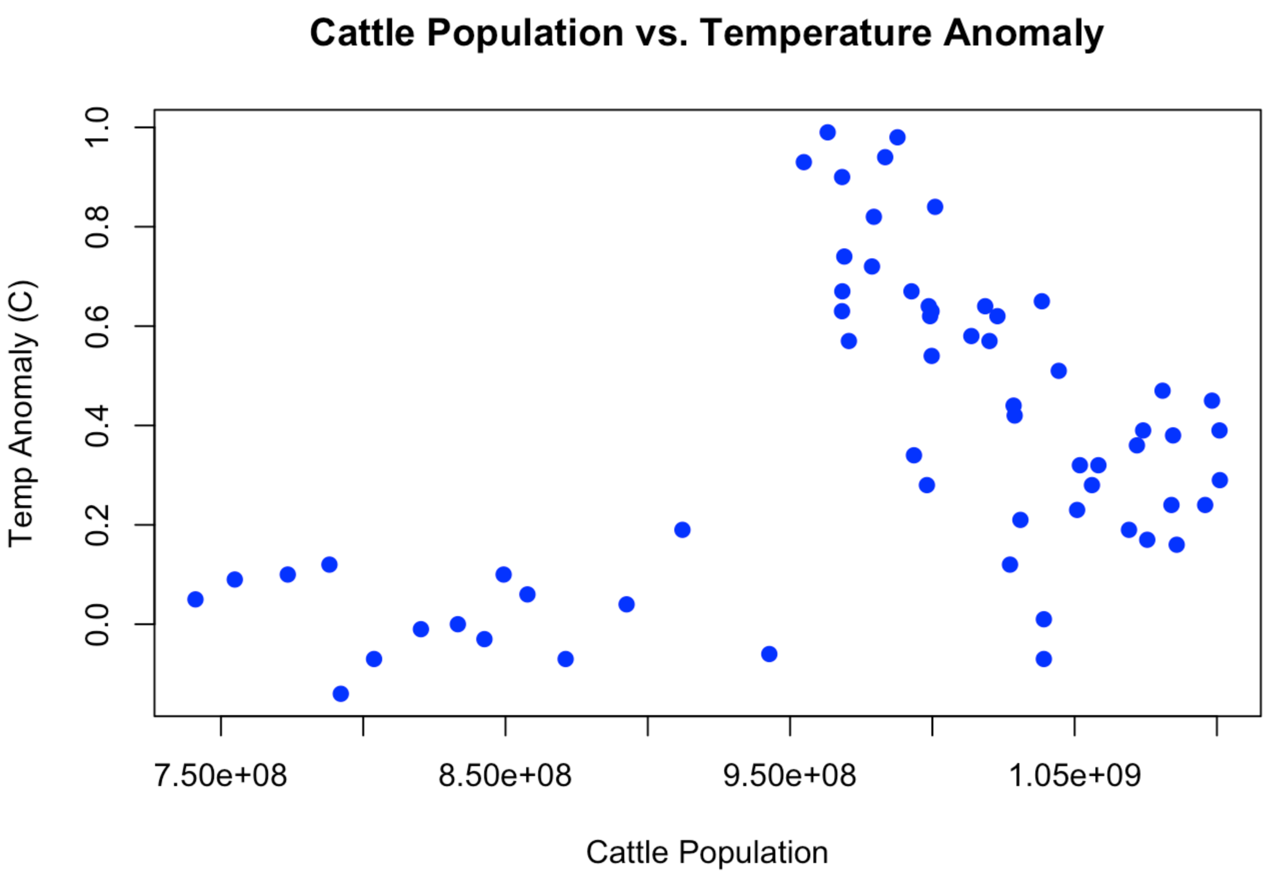
These results tell us that there is a strong linear relationship between CO2 concentration and the temperature anomaly; the p-value for the T-statistic is quite small, indicating that the associated parameter is highly significant.

On the other hand, the cattle population does not appear to have a strong linear relationship with the temperature anomaly. The T-statisitic p-value is quite high, about 90%. This is indicitave of an insignificant coefficient.

The model has a very high R^2 value; 91% of the variation is explained by the model, no doubt due to the strong relationship observed between CO2 concentration and temperature anomaly.

3.2.2. Model B

After revisiting the time series plots in section 2.2, it seems clear that there would not be a strong linear relationship between cattle population and temperature anomaly. Indeed, observe the following scatter plot of the two series:



So the naive linear model will not suffice, at least for the inclusion of the cattle time series.

All the time series indicate some level of trend, so we will incorporate trend into our model. But the trends in each time series do not seem to be simply linear, so we will consider an additional quadratic trend term as well. Thus, our second model will be

$$G_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 CO_t + \beta_4 CA_t + W_t.$$

Below is the summary table of the results after fitting the above model:

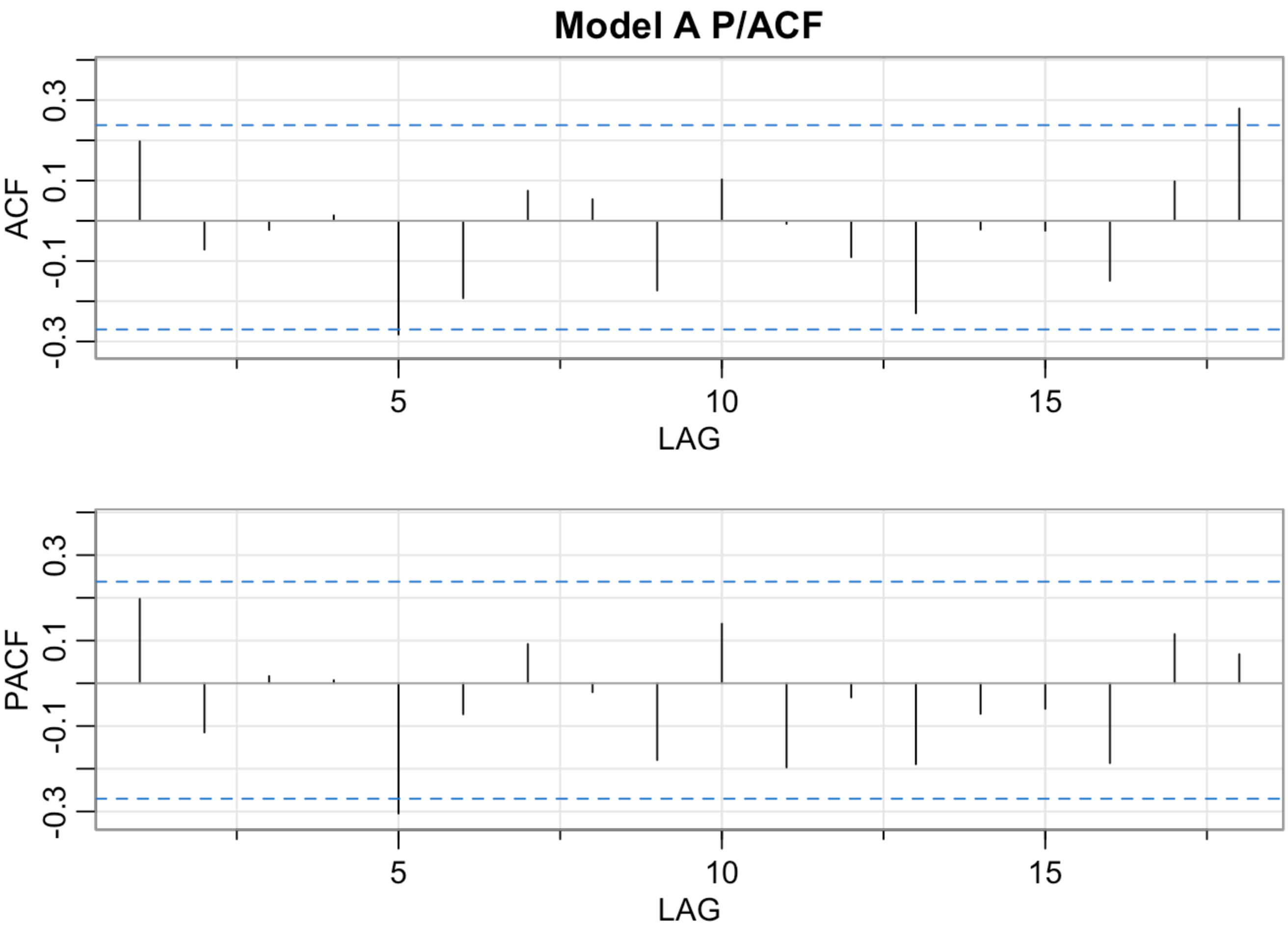
```
##
## Call:
## lm(formula = temp_ts$Anomaly[81:142] ~ I(1:62) + I((1:62)^2) +
##     as.numeric(co2_ts$CO2[3:64]) + cattle)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.146266 -0.059114 -0.007613  0.064113  0.164863
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -2.295e+01  6.003e+00  -3.824 0.000328 ***
## I(1:62)                      -3.981e-02  1.299e-02  -3.064 0.003336 **
## I((1:62)^2)                  -1.003e-03  3.196e-04  -3.139 0.002684 **
## as.numeric(co2_ts$CO2[3:64])  7.422e-02  1.953e-02   3.801 0.000353 ***
## cattle                       -6.836e-10  3.485e-10  -1.962 0.054693 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08628 on 57 degrees of freedom
## Multiple R-squared:  0.9259, Adjusted R-squared:  0.9207
## F-statistic: 178.1 on 4 and 57 DF,  p-value: < 2.2e-16
```

As before, the CO2 series is highly significant. But we also observe that indeed, both trend terms are also significant. And perhaps most interestingly, the incorporation of the trends has "increased the significance" of the cattle population term. Although still technically not significant at a 5% cutoff, this is an indication that the cattle population may play a role in the temperature anomaly.

3.2.3. Model Evaluation and Canonical Model

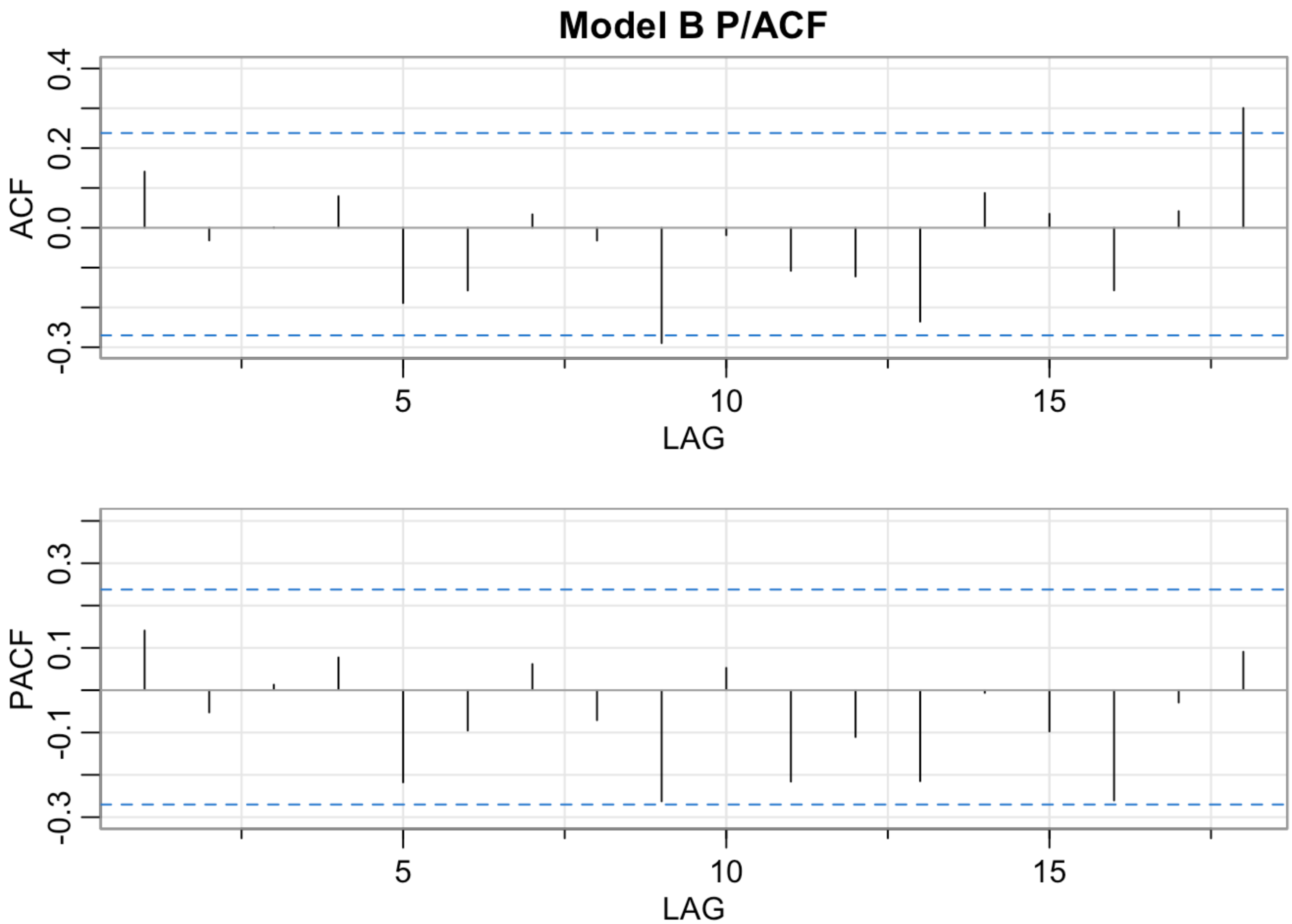
First of all, both models contain no strong evidence of autocorrelation in the errors.

The ACF / PACF of the residuals of model A:




```
##      [,1]  [,2]  [,3] [,4]  [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF   0.2 -0.07 -0.02 0.01 -0.28 -0.19 0.07  0.05 -0.17  0.10 -0.01 -0.09 -0.23
## PACF  0.2 -0.12  0.02 0.01 -0.30 -0.07 0.09 -0.02 -0.18  0.14 -0.20 -0.03 -0.19
##      [,14] [,15] [,16] [,17] [,18]
## ACF  -0.02 -0.02 -0.15  0.10  0.28
## PACF -0.07 -0.06 -0.19  0.12  0.07
```

The ACF / PACF of the residuals of model B:




```
##      [,1]  [,2] [,3] [,4]  [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  0.14 -0.03 0.00 0.08 -0.19 -0.16 0.03 -0.03 -0.29 -0.02 -0.11 -0.12 -0.24
## PACF 0.14 -0.05 0.01 0.08 -0.22 -0.10 0.06 -0.07 -0.26  0.05 -0.22 -0.11 -0.21
##      [,14] [,15] [,16] [,17] [,18]
## ACF   0.09  0.04 -0.16  0.04  0.30
## PACF -0.01 -0.10 -0.26 -0.03  0.09
```

So residual autocorrelation is not an issue with either model.

Model B has a higher R^2 and adjusted R^2 value, but these alone are not sufficient to determine if model B is preferred. We will now examine the AIC and BIC of each model.

Model A has AIC -114.2304677 and BIC -105.7219301, while model B has AIC -121.0834863 and BIC -108.32068. Model B has both a lower AIC and BIC. Therefore, all things considered, model B is the preferred model and will therefore serve as our canonical model for the duration of this report.

4. Results

In this section, we will use the canonical model developed in the previous section.

4.1. Objective 1: Quantitatively Establish Relationship Between CO2, Cattle Population,

and Temperature Anomoly

Although the cattle population is technically not significant, it is close enough to the cutoff that we will consider it for this section; for a more detailed discussion, see section 4.2.

First of all, there is of course a positive correlation between CO2 concentration and temperature anomaly. This isn't a secret, and as mentioned earlier, the purpose of this report was not to uncover this relationship; rather, we hoped to give specific numbers regarding this relationship. And indeed, the model accomplishes just that: for every 1 PPM increase in CO2, the model predicts a 0.07422 degree increase in the temperature anomaly.

Somewhat surprisingly, however, the model assigns negative coefficients to the trend terms as well as the cattle population. Furthermore, the coefficients are qutie small, indicating that the effect of these variables, while significant, is weak.

4.2. Objective 2: Role of the Cattle Population

The role of the total cattle population in the global temperature anomaly is unfortunately still somewhat vague in light of this analysis. The p-value for the associated parameter is insignificant -- albeit barely -- at the 5% level. Not to mention that Model A outright ruled out cattle as a significant predictor quite unambiguously.

These results are inconclusive; we cannot say for certain one way or another whether the cattle population is significant. Further analysis is necessary. Additionally, it is possible that simply looking at the raw population total itself is the wrong way to go about investigating this relationship. Perhaps some other factors should be considered, such as the diet of the cattle, the amount of methane in the atmosphere, whether cattle methane production has been regulated, etc.

4.3. Objective 3: Forecasting Average Global Temperature Anomaly for the Next Five Years

To forecast, we will use the drift method (Hyndman, 2018) to forecast the future predictor values, and then we will use these values in our model to predict the temperature anomaly.

Recall that for predicted values, we have

$$\hat{G}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 CO_t + \hat{\beta}_4 CA_t.$$

Using our model, the forecasted temperature anomalies for the next five years are shown below.

Forecasted temeperature anomalies for the next five years.

Year	Anomaly
2022	0.90
2023	0.85
2024	0.80
2025	0.75
2026	0.69

Some comments: it seems that the predicted temperature anomaly is actually decreasing, contrary to expectations. There are two potential reasons for this. First, it is possible that the quadratic trend term dominates for large t; its coefficient is negative, and t^2 grows faster as t gets larger. This would indicate that a quadratic trend term is not significant in the long term. The other reason is that the model is correct, and it just happens that the next few years are forecasted to have a lower temperature anomaly. This is not unreasonable; the temperature anomaly exhibits seasonal patterns, and there are stretches of several years where the anomaly actually decreased. So we cannot rule out this possibility.

5. Discussion

Below, we offer some comments on the results and how they pertain to our three stated objectives.

Both model A and model B attribute a significant positive effect to CO2 on the temperature anomaly, and so we have successfully quantified the relationship between CO2 and temperature anomaly. However, the role of the trend and cattle are still somewhat vague after the analysis. The forecasted values indicate there is a possibility that the trend terms, particularly the quadratic term, may be inaccurate in the long term.

On that note, the role of the cattle population is also still unknown in light of this analysis. As we previously discussed, further analysis is necessary to ascertain this relationship as there may have been some limitations and red herrings in only considering the raw cattle population.

The forecasted values, if accurate, show that we can expect a decrease in the anomaly for the next few years. Looking at the historical data, this is not unusual. However, if the trend continues, and in particular if CO2 emissions continue unabated, we can expect this anomaly to rebound and continue to increase eventually.

We would also like to briefly comment on the statistical and practical import of the results. We believe that quantifying the relationship is both statistically and practically important; assigning actual numbers to the relationship between temperature and the other variables gives a concrete idea of just how strong the relationship is; awareness of this fact is important and will hopefully encourage environmental conscientiousness. The other results are of questionable import; again, the cattle population needs to be further investigated, and the trend terms may be impacting the forecasting; without knowing for sure, the future forecasted values are not too meaningful.

6. Bibliography

1. *Global Annual Temperature Anomaly* (2022) [Data set]. NOAA. https://www.ncdc.noaa.gov/sotc/global/202113 (https://www.ncdc.noaa.gov/sotc/global/202113)

2. Tans, P., & Keeling, R. (2022). *Mauna Loa CO2 Annual Mean* [Data set]. NOAA/GML, & Scripps Institution of Oceanography. https://gml.noaa.gov/ccgg/trends/data.html (https://gml.noaa.gov/ccgg/trends/data.html)

3. Cook, R. (2022). *World Cattle Inventory by Year* [Data set]. USDA. https://beef2live.com/story-world-cattle-inventory-1960-2014-130-111523$ (https://beef2live.com/story-world-cattle-inventory-1960-2014-130-111523$)

4. Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice* (2nd edition) OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on December 2, 2022.

7. Appendix (Code)

```
# Loading Data

library(astsa)

co2_ts = read.table('co2_annmean_mlo.csv', comment.char = "#", sep = ',', header = FA
LSE, col.names = c('Year','CO2','unc'))

temp_ts = read.table('1880-2021_2.csv', comment.char = "#", sep = ',', header = FALS
E, col.names = c('Year','Anomaly'))

yrs = 1960:2021
cattle = 1000*c(741042,754858,773481,788119,792113,803796,820265,833247,842589,84933
9,857709,871127,892552,912177,942702,1039186,1039180,1030944,1027284,1050868,1056084,
1058348,1069099,1071885,1075512,1085867,1084019,1084584,1100908,1101036,1098269,10740
66,1095866,998084,993533,1080900,1051841,1044419,1038476,1028550,1028924,1020078,1022
868,1018559,1013697,992662,999713,999204,999759,998729,978785,970663,968251,968358,96
9055,954835,963214,968284,979440,983414,987751,1000967)

cattle_ts = data.frame('Year' = yrs, 'Cattle' = cattle)

# Plots

tsplot(x = temp_ts$Year, y = temp_ts$Anomaly, main = "Temperature Anomaly", ylab = "D
egrees C", xlab = "Year")

tsplot(x = co2_ts$Year, y = co2_ts$CO2, main = "CO2 Levels", ylab = "PPM" , xlab = "Y
ear")

tsplot(x = cattle_ts$Year, y = cattle_ts$Cattle, main = "Cattle Population", ylab = "
Number of Cattle", xlab = "Year")

# Model A

fit = lm(temp_ts$Anomaly[81:142] ~ as.numeric(co2_ts$CO2[3:64]) + cattle_ts$Cattle)

summary(fit)

# Temp vs Cattle Scatter plot

plot(x = cattle, y = temp_ts$Anomaly[81:142], xlab = "Cattle Population", ylab = "Tem
p Anomaly (C)", main = "Cattle Population vs. Temperature Anomaly", pch=19, col = 'bl
ue')

# Model B

fit2 = lm(temp_ts$Anomaly[81:142] ~ I(1:62) + I((1:62)^2) + as.numeric(co2_ts$CO2[3:6
4]) + cattle)

summary(fit2)

# P/ACF plots
```

```
acf2(fit$residuals, main = "Model A P/ACF")

acf2(fit2$residuals, main = "Model B P/ACF")

# AIC, BIC

AIC(fit), BIC(fit), AIC(fit2), BIC(fit2)

# Forecasting

co2pred = as.numeric(co2_ts$CO2[64]) +  c(1,2,3,4,5) * (as.numeric(co2_ts$CO2[64]) -
as.numeric(co2_ts$CO2[3]))/61
cattlepred = cattle[62] + c(1,2,3,4,5) * (cattle[62] - cattle[1])/61
timepred = 63:67
timepred2 = timepred^2

new.data = matrix(c(1,1,1,1,1,timepred,timepred2,co2pred,cattlepred),5,5)

preds = round(new.data %*% fit2$coefficients,2)

tabpred = data.frame("Year" = c(2022,2023,2024,2025,2026), "Anomaly" = preds)
knitr::kable(tabpred, caption = "Forecasted temeperature anomalies for the next five
years.")
```